R E P O R T R E S U M E S

ED 013 794 SP 001 306

EFFECTS OF FEEDBACK AND PRACTICE CONDITIONS ON THE ACQUISITION OF A TEACHING STRATEGY.

BY- ALLEN, DWIGHT W. AND OTHERS

PUB DATE 66

EDRS PRICE MF-$0.25 HC-$0.84 21P.

DESCRIPTORS- *DISCRIMINATION LEARNING, *FEEDBACK, QUESTIONING TECHNIQUES, RETENTION, ROLE PLAYING, STUDENT REACTION, *TEACHER INTERNS, *TEACHER SUPERVISION, *TEACHING TECHNIQUES, THOUGHT PROCESSES, VIDEO TAPE RECORDINGS

TO COMPARE SEVERAL METHODS OF DEVELOPING CLASSROOM QUESTIONING (PROBING) TECHNIQUES VIA DISTRIBUTED PRACTICE AND IMMEDIATE FEEDBACK, WHEN THE LATTER EMPLOYED VIDEOTAPED PERFORMANCES OF THE LEARNER, 85 INTERNS WERE VIDEOTAPED ON 4 OCCASIONS DURING THE FIRST 20 MINUTES OF REGULAR CLASSROOM LESSONS. IN BETWEEN TAPINGS THEY RECEIVED 30 MINUTES OF SUPERVISION, IN WHICH THEY VIEWED PLAYBACKS OF EARLIER TEACHING ALONG WITH A CRITIQUE FROM AN EXPERIMENTER WHO PROVIDED DISCRIMINATION TRAINING. WITHIN-SESSION FEEDBACK WAS HELD CONSTANT, AND AMOUNT OF PRACTICE AND DELAYED FEEDBACK WAS MANIPULATED, OVER 4 EXPERIMENTAL GROUPS. A POST-TEST WAS VIDEOTAPED ABOUT 7 WEEKS AFTER TRAINING. INTERNS WERE TRAINED IN PROBING TECHNIQUES (CLARIFICATION, CRITICAL AWARENESS, REDIRECTION, PROMPTING, REFOCUS) WHICH DEPENDED ON PUPIL RESPONSE, AS WELL AS AN ENCOURAGING DIVERGENT THINKING, ROLE PLAY IN BRIEF, AND PUPIL SUMMARY. TREATMENT DIFFERENCES, THOUGH NOT ENTIRELY CONSISTENT, FAVORED MASSED PRACTICE-IMMEDIATE FEEDBACK OVER DISTRIBUTED PRACTICE-REINSTATED FEEDBACK IN INITIAL ACQUISITION OF PROBING BEHAVIORS. THE FORMER ALSO PRODUCED SIGNIFICANTLY MORE FREQUENT PROBING THAN DISTRIBUTED PRACTICE AND IMMEDIATE FEEDBACK. RETENTION INFERENCES CAN BE DRAWN FROM THE FACT THAT DISTRIBUTED PRACTICE-DELAYED FEEDBACK GROUPS MAINTAINED HIGHER PROBING RESPONSE RATES ON THE POST-TEST THAN DID MASSED PRACTICE-IMMEDIATE FEEDBACK. (AF)

SEP 1 8 1967

SP 01306

R 6 od

ED013794

~~Experiment II:~~

EFFECTS OF FEEDBACK AND PRACTICE CONDITIONS ON THE ACQUISITION OF A TEACHING STRATEGY

Dwight W. Allen, Frederick J. McDonald, and Michael E. J. Orme
(Stanford University)

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

EFFECTS OF FEEDBACK AND PRACTICE CONDITIONS ON THE ACQUISITION OF A TEACHING STRATEGY

A key problem in the development of instructional systems designed to produce teaching skills is that of providing adequate feedback on the teacher's performance. Quite apart from the practical limitations encountered in trying to schedule immediate feedback sessions with teacher trainees, supervision tends to focus on the end products of performance rather than the course of learning. This is simply due to the fact that it is not usually possible to provide feedback during the teaching performance. Discrimination training is necessarily based on the perceptions of the teacher and the supervisor as they remember the lesson. The entire process thus tends to invite a heavy reliance on private frames of reference. The supervisor and the teacher do not start from a common perception of what was done, how it was done, and what the effects were.

Perhaps the signal advantage of televising trainee lessons is that in subsequent supervision sessions, the original performance can be completely reinstated. In this way, the teacher intern is not required to respond to supervision on the basis of what he and the supervisor recall about a complex series of events. Further, the usual inhibitory effects of delayed feedback may be offset, not only because the initial teaching performance is reinstated, but because it is possible to provide discrimination training at any point in the development of a response sequence. Rather than reacting to performance in terms of end products, it is possible for the intern and supervisor to analyze a given interaction sequence in the classroom, identify salient cues, and develop strategies to improve further performance. The results of an earlier experiment by McDonald, Allen, and Orme (1966) support these hypotheses. They found that the provision of discrimination training during videotaped playbacks of intern lessons produced significantly greater increases in selected teacher behaviors than did confirmation or self-feedback procedures.

If the above reasoning is valid, and the reinstatement of original performance by videotape does in fact solve the problem produced by delaying feedback, the problems concerning the optimal arrangement of feedback and interspersed practice sessions take on new meaning. For example, given the capability to reinstate the original performance by videotape at any point in time, should practice precede or follow supervision? In the first case, the acquisition of a complex skill might be slower. However, substantial gains might be

realized because the learner becomes less defensive about feedback based on his earlier performance, and thus becomes more responsive to supervision. This would be "delayed" feedback in the sense that the learner was receiving training based on an earlier performance. In this kind of a training situation, the learner would be serving as his own model.

Where supervision immediately followed performance, the learner's ego-involvement in the lesson just taught might lead him to be less open to suggested change. It is also probable, of course, that the immediacy of supervision would outweigh such potential resistance.

There is a general consensus that initially at least, massed practice is optimal for the acquisition of complex skills and reasoning strategies. Given time limitations in training, however, we are still left with the problem of retention. McDonald, Allen, and Orme (1966) found that while interns quickly reached criterion (where the dependent variable was amount of reinforcing of pupil participation) initial gains were not lasting. This result suggests that distributed practice would be a logical alternative. Further, given the perceptual adequacy of videotaped feedback, it is quite possible that acquisition rates would not be significantly lower than under massed practice conditions.

Finally, while there is considerable evidence to suggest that distributed practice is generally superior to massed practice, why this is so is not known. Recent research by McCaugh and Hostetter (1961) tends to support a consolidation theory of learning. However, as Hilgard (1962) points out, none of the current theories account for exceptions to the general rule. In training, then, the optimal organization of practice periods becomes an inelegant process of empirically establishing the limits of each approach in terms of a given skill or strategy. The purpose of the experiment reported here was to compare several methods of distributing practice and feedback when the latter employed videotaped performances of the learner.

General Procedure: The general procedures followed in this research were similar to those reported in the first experiment (McDonald, Allen, and Orme, 1966) in a series of feeback studies. Intern teachers were videotaped on four separate occasions, during the first 20 minutes of regular classroom lessons. In the intervals between each of the taping sessions, they received experimental treatments within the context of regular supervision.

The treatment of supervision sessions were alike for all subjects in that they viewed videotaped playbacks of their earlier teaching performance with an experimenter who provided discrimination training. However, instead of varying the amount of feedback within each session as in the first experiment, we held within-session feedback constant, and manipulated the amount of practice and delayed feedback over four experimental groups. In addition, the post-test was videotaped approximately seven weeks after training so as to permit inferences about retention as well as acquisition curves.

The Dependent Variable: In the earlier experiment, interns had been trained to reinforce pupil participation in discussion and review lessons. A natural extension of this skill seemed to be those techniques where the teacher could increase the quality of such participation.

Rather than attempting to operationally define a "penetrating question," or to develop techniques for suppressing superficial "first-answer" pupil responses, the approach taken was to develop classroom techniques which followed simple shaping procedures. Interns were given three basic rules: (1) Do not give immediate answers to pupil questions. (2) Once a pupil has responded, try to get him to "go beyond" the information given (by one of several specified techniques). (3) Differentially reinforce pupil responses that demonstrate increased critical awareness.

In the initial written instructions to each intern, and in supervision as well, a series of discrete techniques were presented following the statement of each rule. The basic strategy in each case was the same. Following a pupil response, the teacher asked a question designed to elicit more information or more meaning from the pupil. If the pupil response was adequate or insightful, the teacher then attempted to get him to relate the questions, answer, or comment to another area, or to spell out its implications for a given issue.

These various techniques were presented as exemplars of a basic questioning strategy termed <u>probing</u>. Specific techniques were defined as: <u>clarification</u>, <u>critical awareness</u>, <u>redirection</u>, <u>prompting</u> and <u>refocus</u>. (The term used to describe each specific category reflects the teacher's goal when using a given technique.)

In addition, a secondary set of techniques designed to achieve the same ends as probing procedures were included as part of the training problem.

The latter techniques differ from probing in that they may or may not depend on a prior pupil response, i.e., the teacher can introduce them at any point in the discussion himself, or use them as specific types of probes. These techniques were termed encouraging alternative (divergent thinking), supposition (role play in brief) and pupil summary of discussion. Both in training and in later tape analysis, the pupil response served as an S^d to cue the intern or rater that a teacher probing (classifiable) response should immediately follow.

Treatment Conditions: Stanford interns were assigned to one of four experimental groups, each of which received varying amounts of practice between feedback sessions. Subjects in each group received discrimination training from one of two Es who were present at each feedback session. All feedback sessions were thirty minutes in length. A summary of the design appears in Table I. (See Table 1, Experiment II).

Group 1. (Immediate Feedback --Massed Practice): Group 1 subjects, like those in the other three groups, initially received written instructions on probing. The instruction related to new techniques to the interns' prior training in reinforcement, and stated that the basic aim of the current study was to help interns develop a broader range of basic questioning skills in discussion and review lessons. Each rule was then stated, and specific techniques together with relevant examples were presented. One of the two E^s then briefly reviewed the techniques with the intern, reinforced positive statements about the potential utility of probing techniques, and viewed the intern's playback with him.

Each time a pupil responded verbally, E cued the intern, and if the latter had probed then E reinforced him. If he had not, E suggested how this might be done. Interns in the first group were both videotaped and supervised on three consecutive days. They were videotaped at the school on a given day, and received supervision based on that performance the same evening. The post-test was videotaped 45 days after the pretest was made.

Group 2 (Immediate Feedback - Distributed Practice): These subjects received the same kind of discrimination training as those in Group 1. Instead of receiving massed practice however, they were directed to practice probing techniques for a one-week period following each of the supervision sessions.

Supervision sessions were based on videotapes that had been recorded earlier on the same day. The Post-test followed the pretest videotape by 45 days.

Group 3 (Delayed Feedback -- Distributed Practice): Subjects in this group received the same supervision treatment as those in Group 2. However, the taping schedule was arranged in such a way that they always viewed a videotape of their performance which had been filmed one week earlier. This has been termed delayed feedback since unlike Group 1 and 2 subjects who received feedback based on that day's performance, these subjects were taped on the day following supervision, and received no discrimination training based on their performance for one week. At the conclusion of each session Group 3 subjects were directed to practice probing techniques not only in the next day's lesson, but during the rest of the week as well.

Group 4 (Reinstated Feedback - Distributed Practice): These subjects received treatment distributed over a six-week period. Following taping on the first day of week one, they received supervision based on that tape on the first day of week two. Throughout treatment, they were taped and supervised on alternate weeks. As was the case with the first three groups, the fourth group were post-tested 45 days after the pretest had been filmed.

Subjects: Interns were selected from the same population as that for the first experiment.[1] A total of 85 intern teachers majoring in English, Social Studies, Mathematics or Science were selected for study. Approximately equal numbers of interns from each subject-matter area, and from socio-economically equivalent schools were assigned to each of the four groups. Mean age for the four groups varied from 23.7 to 25.5 years. Sex differences and subject-matter major in each group were very similar to those already described in the first experiment.

Measurement procedures: Four videotapes for each intern in each group were analyzed by raters trained for this purpose. Four raters received intensive training on non-experimental tapes before the analysis of the latter was begun. Training consisted of joint rating sessions that continued until there was 95% agreement on the major response categories.

Reliability was maintained by frequent analyses of double rated tapes. In addition, joint rating sessions were held after each block of 50 tapes had been completed so that systematic rater biases could be forestalled. As rating progressed, it was found that certain tapes inevitably yielded low interrater agreement. This occurred when the intern in question was unable to maintain classroom discipline, or when there was an unusually high

---

[1] There was a three-month break between the conclusion of the first study and the beginning of the study being described here.

rate of interaction combined with "fuzzy" audio or visual output. In these cases, the tapes were rated by independent teams of raters -- by dividing the rating task between themselves and replaying frequently, two raters were able to record all interactions. To control for combined rating effects, independent teams were used. The identification of these "trouble" tapes posed a problem, since when raters were asked to rank the rating difficulty of tapes their judgments were not entirely consistent with reliability checks. For this reason, it was finally decided to double-rate all tapes.

The results of these control procedures are reflected in the reliability coefficients reported in Table 2. (See Table 2, Experiment II) The videotapes on which they are based were selected in the following manner. Since coefficients based on all of the tapes would have been prohibitively expensive and also time-consuming, 160 tapes were selected using tables of random numbers. Five tapes from each group for each trial were randomly selected from pools of 15 to 21 tapes.[2] The coefficients reported then are based on a representative selection of independently rated tapes from each group and for each trial. As can be seen, interrater agreement on the major response categories is very high.

A certain amount of data were lost between initial videotaping and statistical analysis. Forty-four tapes were omitted at the outset because of technical inadequacy or unduly short tape time. Fifty-six tapes were prorated to bring them up to the 20.0 minute criterion. Tapes less than 15 minutes in length were omitted, those between 15 and 20 minutes were prorated. Before any of the above adjustments were made, the T statistic was applied to mean tape time and number of omitted tapes per cell to determine whether or not any significant differences between cells existed. The results were well short of significance as these tapes were almost equally divided among all cells. The standard errors for short tapes varied from .40 to 1.2 minutes. As a further check on the data, the mean number of days between pre and post-test were calculated for each experimental group. The overall mean was 45.25 days, with $S_e = 1.41$ days.

RESULTS

Probing was presented as a basic questioning strategy in training, and specific techniques such as Refocus or Critical awareness were subsumed under this broad general concept. It is important to note, however, that this was a logical distinction made to

---

[2] Team ratings were not included in selection because they had been rated by special techniques, and were known to be highly reliable on all major response categories.

facilitate training. It does not follow that it is psychometrically relevant. In fact, the analysis of results did not proceed in terms of one broad dependent variable, but in terms of eight. Each of the questioning and related techniques were viewed as discrete dependent variables. Table 3 provides support for this procedure. The intercorrelations between the subcategories of probing and non-probing are very low. Out of the 78 relevant coefficients, only 5 are significantly different from zero. Of these, 3 reach the .05 level, and 2 reach the .01 level. Prompting is significantly related to redirect ($r = .23$, $p = .05$), refocus ($r = .28$, $p = .05$) and clarification ($r = .32$, $p = .01$).

However, prompting occurred infrequently (group means varied from 3.7 to 1.7) and contributed little to overall differences.

In the analysis of treatment differences then, F Ratios for all of the major probing techniques must be considered. F ratios based on probing provide a general overview of the results taken as a whole, but since they include variables which did not reflect significant differences, there is a suppression effect.

Treatment Differences: Analyses of covariance with relevant trial 1 scores as covariates were carried out to determine between-group differences in trials 2, 3, and 4. Table 4 summarizes these results. As can be seen, differences are reflected in several of the dependent variables.[1]

Treatment differences were most prominent in trial 3. The experimental groups were found to differ significantly from each other in the use of critical awareness ($p < .01$), redirection ($p < .05$), refocus ($p < .01$) and role play ($p < .05$) however, the relative frequency of occurrance for each of the response categories that determine treatment differences is important in attaching meaning to these results. For example, role play occurs infrequently and was significant, while clarification which contributes heavily to total teacher responses was not. This is reflected in the rather low F ratio for probing, which while it suggested a strong trend, was not significant ($p < .250$).

Treatment differences in trials 2 and 4 are much less general. In trial 1, redirection was significant ($p < .05$) and clarification, refocus and pupil summary approach significance, but within-group variance is great enough to reduce the F ratio to 0.600, for probing. By trial 4 only refocus shows a clear trend ($p < .10$) for treatment effects.

---

[1] In general, the most frequently occurring responses are listed first in the tables. Further, an inspection of Table 3 shows that the more frequently a response occurs, the more highly correlated it is with probing.

An overall summary of treatment differences derived from the analyses of covariance was shown in Table 5. (See Table 5, Experiment II) Adjusted means, standard errors and F ratios for mean frequencies of probes by groups over trials are presented. Figures 1 and 2 (See Figures 1 and 2, Experiment II) illustrate the general pattern of probing, and clarify specific between-group differences for a given trial. Both figures are based on unadjusted means so that trial 1 levels of performance can be seen in relation to the other three trials.

In Figure 1, all unadjusted probing means were brought to a common point somewhat analogous to a covariance adjustment by dividing the respective means for trial one into each of the four means for each group. This clearly illustrates specific between-group differences which also appear in Table 6. (See Tabel 6, Experiment II)

In trial 2,Groups 1 and 4 differed significantly ($p < .05$) from Group 2. Group 3 subjects also tended ($p < .10$) to redirect more frequently than did Group 2 subjects. In trial 3, Groups 1 and 3 interns used redirection, refocus and critical awareness significantly more often than did Group 2 and Group 4 subjects (Table 6).

Figure 1 tends to exaggerate certain group trends, particularly from trials 3 to 4. Figure 2 is singularly instructive here because it considers all of the dependent variable in relation to total teacher responses. In this figure plot points were derived by expressing probes as a percentage of total teacher responses. Now, comparing figure 1 with figure 2, it can be seen that Group 3 and 4 do not drop off from trial 3 to trial 4 as implied in figure 1. The relevant conclusion then is that the major treatment differences still lie in trial 3, but that retention rates for Groups 2, 3, and 4 are considerably higher than a consideration of probes in isolation to other responses would indicate.

Training Differences: Table 7 (See Table 7, Experiment II) summarizes the training or within-group differences for the four groups. The probing category, effectively represents these differences. Group 1 subjects realized their greatest gains in the discrete probing techniques by trial 3. This was also true for Group 3 and 4 subjects. However, as Figure 2 illustrates, the latter groups did not drop off as extensively as did Group 1 in later trials.

Discussion of Results

Treatment differences were not entirely consistent throughout trials and across dependent variables, so that conclusions must be qualified. It is clear however that there were significant treatment differences in favor of Group 1 (Massed Practice and Immediate feedback) over Group 4 (Distributed Practice and Reinstated Feedback) in the initial acquisition of probing behaviors. This occurred by trial 3.

Group subjects also probed significantly more frequently than Group 2 (Distributed Practice and Immediate Feedback) subjects, although differences between these latter two groups may also be influenced to some extent by differences in the temporal contiguity between supervision and subsequent videotaping. Group 1 subjects were taped immediately after supervision, while Group 2 subjects were exposed to a one week practice period before they were videotaped the next time.

Certain inferences can also be drawn about treatment differences related to retention. Figure 2 indicates that the distributed practice and delayed feedback groups (Groups 2 and 4) maintained relatively higher probing response rates on the post-test than did Group 1 who dropped off quite sharply. However, the data do not permit firm conclusions here as F ratios for trial 4 were not significant.

Finally, it should be pointed out that there appear to be practical limits on the amount of probing possible in any given class period. Unlike teacher reinforcement which may occur in high frequency, classroom and subject matter concerns establish a ceiling for probing, and this is why the most sensitive measures of this variable are likely to be expressed as ratios or percentages related to total teacher responses.

TABLE 1
(Exp. II)

SUMMARY OF THE EXPERIMENTAL DESIGN

Practice Conditions	Feedback on Teaching Performance	
	Immediate	Delayed
Massed Practice	Group 1 (N = 21)	-----
Distributed Practice (1 Week Intervals)	Group 2 (N = 21)	Group 3 (N = 21)
Distributed Practice (2 Week Intervals)	-------	Group 4 (N = 22)

TABLE 2
(Exp. II)

RELIABILITY COEFFICIENTS BASED ON A RANDOMIZED SAMPLE OF 80 VIDEOTAPES RATED BY TWO INDEPENDENT OBSERVERS

Response Category	Reliability Coefficient
Total Pupil Responses	0.9934
Total Teacher Probes	0.9939
Total Teacher Non Probes	0.9871
Total Teacher Reinforcement	0.9579
Probing Sub-Categories:	
a. Clarification	0.9689
b. Critical Awareness	0.8658
c. Redirect	0.7633
d. Prompting	0.5914
e. Refocus	0.7312

TABLE 3 (Exp. II)

CORRELATION MATRIX FOR THE MAJOR RESPONSE CATEGORIES
OF THE DEPENDENT VARIABLE

N = 79

ependent iables	NPr	AQ	TNR	Pr	Cl	CA	Rd	Pmt	Rf	EA	PS	RP
	19	20	21	22	23	24	25	26	27	28	29	30
Probes Pr) 19	1.00	.4784	.6889	.0969	.0542	-.0246	.1925	.1265	-.0012	.0554	.3089	.1360
swers 20 estions (AQ)		1.00	.4726	-.0803	-.1199	-.0599	-.0046	.1614	-.0333	-.0697	.1673	.1064
No Res- nse (TNR) 21			1.00	.1780	.1197	.0399	.0563	.2121	.1070	.0678	.1576	-.0080
bes (Pr) 22				1.00	.9121	.5072	.5608	.4557	.4027	.1099	.0380	.0858
arifies 1) 23					1.00	.2385	.3269	.3697	.2126	.0233	.0251	.0476
itical Aware- ss (CA) 24						1.00	.1796	.1728	.1338	.0584	.0591	.1313
direct d) 25							1.00	.2416	.2049	.1096	.1351	.0292
ompt t) 26								1.00	.2834	-.0359	-.0016	.0491
focus f) 27									1.00	.0349	.1223	.0855
couraging ternatives A) 28										1.00	.0214	.1569
mmary S) 29											1.00	.0212
le Play P) 30												1.00

OTE: In order for any of the coefficients reported above to be considered significantly different from o, r must = .232 (p = .05) or .303 (p = .01).

N. of Correlation Coefficients = 78

TABLE 4
(Exp. II)

SUMMARY OF F RATIOS AND SIGNFICANCE LEVELS FOR THE DEPENDENT VARIABLES, DERIVED FROM TRIAL TWO, THREE AND FOUR ANALYSES OF COVARIANCE, WITH TRIAL ONE SCORES AS COVARIANTS

Response Category	Trial 2		Trial 3		Trial 4	
	F Ratio	df	F Ratio	df	F Ratio	df
Probing	0.600	3/58	1.416[a]	3/69	0.010	3/48
Clarification	1.217	3/58	0.386	3/69	0.900	3/48
Critical Awareness	0.898	3/58	4.898**	3/69	0.243	3/48
Redirection	3.496*	3/58	3.385*	3/69	0.140	3/48
Prompting	0.092	3/58	0.767	3/69	0.603	3/48
Refocus	1.095	3/58	4.668**	3/69	2.474[b]	3/48
Encouraging Alternatives	0.373	3/58	1.952[a]	3/69	0.463	3/48
Pupil Summary	1.010	3/58	0.807	3/69	0.803	3/48
Role Play	0.595	3/58	3.670*	3/69	1.389	3/48
	N = 63		N = 74		N = 53	

Levels of Significance:

a for F = 1.41, p = .25

b for F = 2.18, p = .10

*for F = 2.76, p = .05

**for F = 4.13, p = .01

TABLE 5
(Exp. II)

SUMMARY OF THE ANALYSES OF COVARIANCE FOR THE EXPERIMENTAL GROUPS, WITH TRIAL ONE SCORES AS COVARIANTS AND RPOBING AS THE DEPENDENT VARIABLE

		Group 1	Group 2	Group 3	Group 4	df	F
		Adjusted Means and Standard Errors					
Trial 2	M	40.88	36.42	43.36	39.76	3/58	0.600
	Se	3.9100	3.5857	3.9128	4.5249		
Trial 3	M	36.08	29.41	37.19	32.12	3/69	1.416[a]
	Se	3.3668	3.2668	3.2605	3.3443		
Trial 4	M	35.93	35.63	35.54	36.31	3/48	0.010
	Se	3.7769	3.5293	4.0819	4.2659		

Level of significance for a: F (3/69) = 1.41, p = .250

TABLE 6
(Exp. II)

TREATMENT DIFFERENCES BETWEEN GROUPS IN TRIAL TWO AND THREE

BASED ON THE T STATISTIC

Response Category	Trial 2		Trial 3	
	Direction of Difference Between Groups Within a Given Trial			
	F Ratio		F Ratio	
Redirection	3.496	$G_2 < G_1$ ($p < .05$) $G_2 < G_3$ ($p < .10$) $G_2 < G_4$ ($p < .05$)	3.385	$G_2 < G_1$ ($p < .10$) $G_4 < G_1$ ($p < .05$) $G_4 < G_3$ ($p < .10$) $G_2 < G_3$ ($p < .10$)
Critical Awareness		NS	4.898	$G_2 < G_3$ ($p < .05$) $G_4 < G_3$ ($p < .01$) $G_4 < G_1$ ($p < .10$)
Refocus			4.668	$G_4 < G_1$ ($p < .01$) $G_2 < G_1$ ($p < .10$) $G_4 < G_3$ ($p < .05$)
Role Play			3.670	Mean frequencies too low to make differences meaningful

TABLE 7
(Exp. II)

TRAINING DIFFERENCES FOR THE FOUR EXPERIMENTAL GROUPS FROM TRIAL TO TRIAL, ON MAJOR RESPONSE CATEGORIES

Response Category	Group 1	Group 2	Group 3	Group 4
	Direction of Difference from Trial to Trial and p level (two-tailed)			
Total Pupil Responses	NS	NS	$T_1 < T_2$ (.05)	NS
Teacher No Response	NS	NS	$T_1 < T_4$ (.10)	NS
Probing	$T_1 < T_2$ (.05) $T_1 < T_3$ (.05)	NS	$T_1 < T_2$ (.05)	$T_1 < T_2$ (.05)
Clarification	$T_1 < T_2$ (.10)	NS	NS	$T_1 < TT_3$ (.10)
Critical Awareness	NS	NS	NS	$T_1 < T_2$ (.05)
Redirection	$T_1 < T_2$ (.10)	NS	NS	NS
Refocus	$T_1 < T_3$ (.05) $T_1 < T_4$ (.05)	NS	NS	$T_1 < T_2$ (.05)
Encourages Alternatives	$T_1 < T_2$ (.10)	NS	NS	NS
Role Play	NS	NS	$T_1 < T_2$ (10)	

(Exp. II)

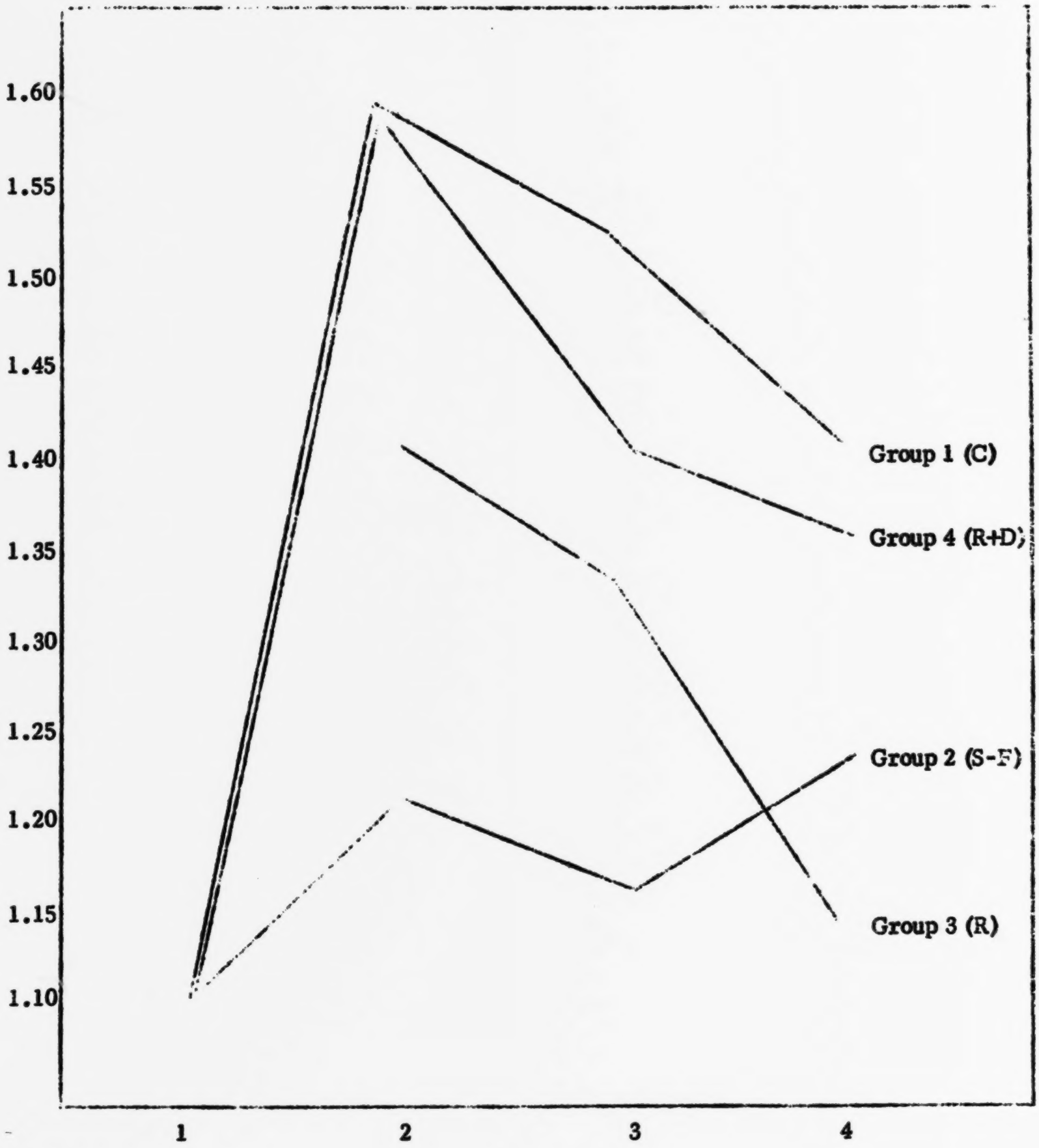

FIGURE 1. Unadjusted mean frequencies for probing brought to a common point by using trial one means as divisions in each trial.

(Exp. II)

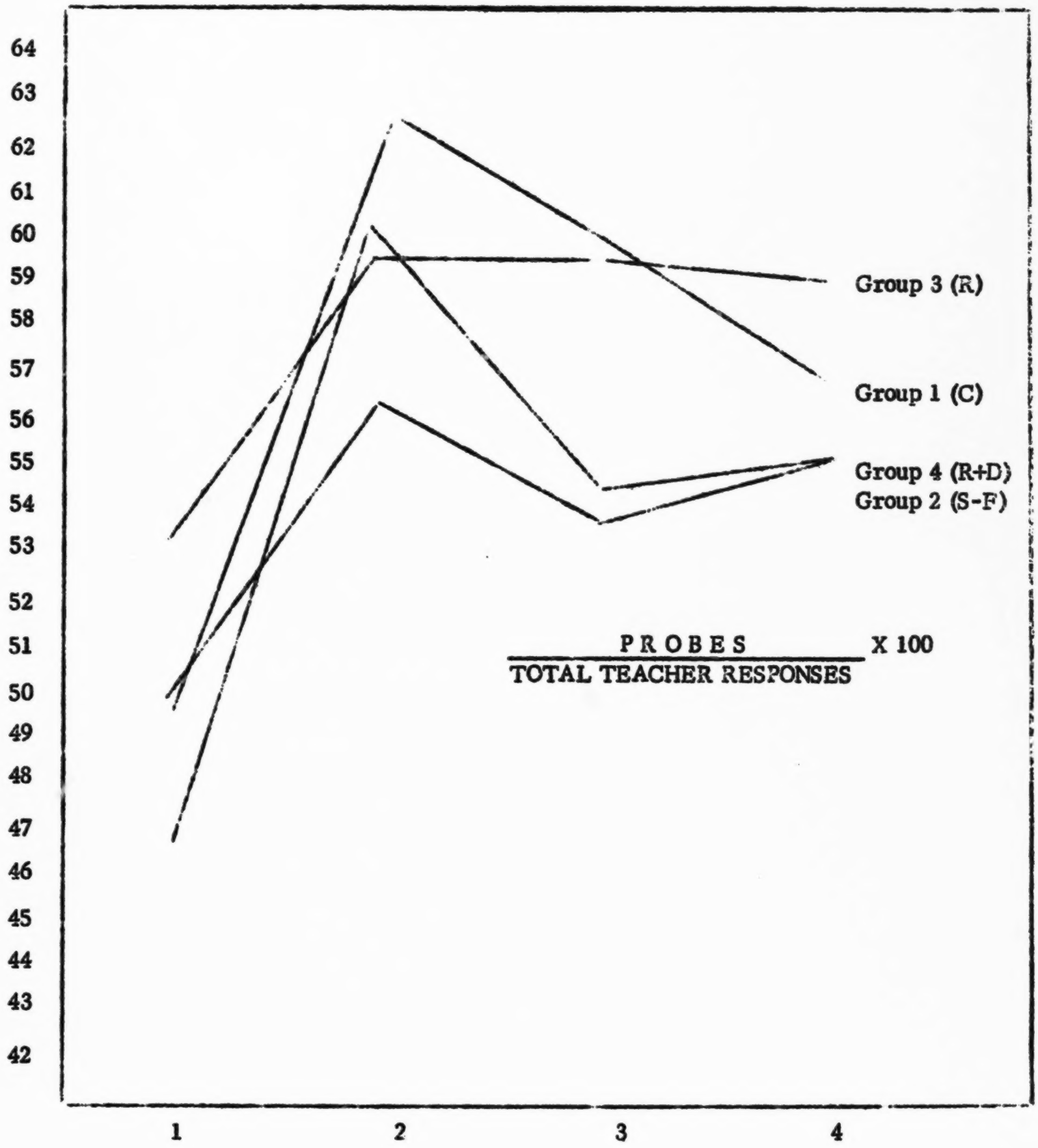

FIGURE 2. Percent of probes contributing to total teacher responses for all groups over all trials.

(Exp. II)

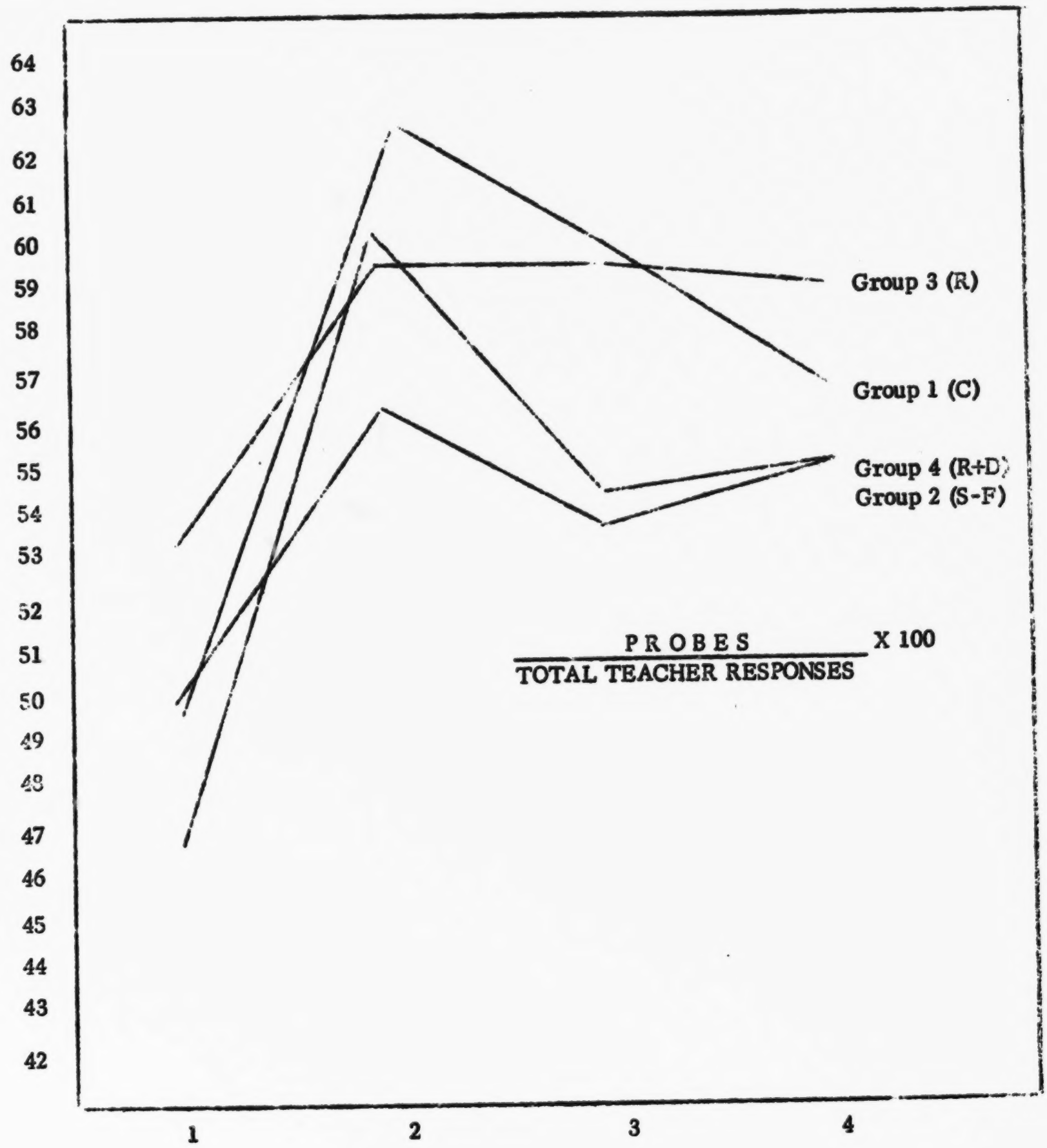

FIGURE 2. Percent of probes contributing to total teacher responses for all groups over all trials.